Postal traffic in Portugal
Applying time series modeling

27TH CONFERENCE ON POSTAL AND DELIVERY ECONOMICS
24 May 2019
Dublin

Filipa Silva
Carla Machado
ANACOM | Portugal

Table of contents

1. Brief context
2. Data source and samples
3. Postal traffic evolution and determinants
4. Methodological framework
5. Estimation, goodness-of-fit and forecast
6. Conclusions
1. Brief context

Past: GDP - the main driver

Evolution of letter-post volume per capita (domestic and international combined) and GDP per capita in industrialized countries (IC) and developing countries (DC) (index 1980 = 100)

Source: UPU Postal Statistics
Present: Digitalization rules


Economic growth

Connectivity
Rate of Internet penetration
Digitalization (e-commerce, e-government, e-substitution)

Trade
Exports

Demand for postal service

Internet affects:
consumers behaviours and
production and distribution
of goods and services

The authors, based in information from UPU (2018).

Present: Digitalization rules


Technology
Digitalization
E-commerce
E-government
E-substitution

Convergence
Trade off between
electronic and
physical services

Demographics
Aging population
Young people
ever more digital

Enviroment
Substitution
Last mile delivery

The authors, based in information from ERGP (2019).
2. Data source and samples

Data source

**Source:** Data from ANACOM, collected from postal services providers

Traffic (domestic and outgoing international), by type of item:

- **Correspondence**
  - Letters, editorial mail and direct mail
- **Parcels**

**Time span:** quarterly data, from 1Q 2005 to 2Q 2018
3. Traffic evolution and determinants

Postal traffic (domestic and outgoing international)

Figure 3 – Evolution of the total postal traffic (domestic and outgoing international), Portugal

Law no. 17/2012, of 26 April: (transposition of the European Directive)
Revised definitions of postal services
Removed the maximum weight limit for postal parcels (previously, 20 Kgs)

2014
New object types reclassification reported by some postal operators (affects parcels, mostly)

Unit: Millions of objects.
Source: Authors, with data from ANACOM.
... by types of items

Letters, editorial and direct mail (94%)

Parcels (6%)

Explanatory variables

Table 1 – Variables considered that may affect postal traffic

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unemployment</td>
<td>Total Unemployment</td>
<td>INE, Portugal</td>
</tr>
<tr>
<td>ESI_Index</td>
<td>Economic sentiment indicator</td>
<td>INE, Portugal</td>
</tr>
<tr>
<td>CCI_Index</td>
<td>Consumer confidence indicator</td>
<td>INE, Portugal</td>
</tr>
<tr>
<td>CPI</td>
<td>Consumer price index (12-month average growth rate)</td>
<td>INE, Portugal</td>
</tr>
<tr>
<td>GDP</td>
<td>Gross domestic product</td>
<td>INE, Portugal</td>
</tr>
<tr>
<td>Exports</td>
<td>Exports of goods and services</td>
<td>INE, Portugal</td>
</tr>
<tr>
<td>Penetration_BB</td>
<td>Fixed broadband accesses, per 100 inhabitants</td>
<td>ANACOM</td>
</tr>
</tbody>
</table>
Correspondence

Figure 6 – Correlations between correspondence traffic and exogenous variables (1Q2005 – 3Q2010).

Parcels

Figure 9 – Correlations between parcel traffic and exogenous variables (1Q2005 – 3Q2010).

Figure 7 – Correlation between correspondence traffic (letters, editorial mail and direct mail) and GDP, exports and fixed broadband penetration, between the 1Q2005 and the date mentioned in horizontal line, Portugal.

Figure 10 – Correlation between parcels traffic and GDP and Internet penetration, between the 1Q2005 and the date mentioned in horizontal line, Portugal.

Sources: INE (Unemployment – total; Unemployment – SSA index – Economic sentiment indicator; CGI index – Consumer confidence indicator; CPI – Consumer price index (12-month average growth rate); GDP – Gross domestic product; Exports – Exports of goods and services) and ANACOM (Penetration, BD – Access to fixed broadband per 100 inhabitants).

5/29/2019
4. Methodological framework

Methodological approaches

(1) ARIMA models

(2) Decomposition models: exponential smoothing

(3) Multiple linear regressions

The best model
ARIMA MODELS

Iterative process

Data preparation
- Transformations the data
- Stationarity

Identification
- Orders of the ARIMA model: ACF and PACF
- ARIMA(p,d,q) or SARIMA \( (p,d,q)(P,D,Q)s \)

Estimation
- The estimation of the parameters can be done by the conditional least squares and maximum likelihood.

Diagnosis tests
- Analyzing the residuals of the estimation.
- Different ARIMAs can be estimated.

Compare the results after estimation

<table>
<thead>
<tr>
<th>Error measures in the estimation period</th>
<th>Residual diagnostics and goodness-of-fit tests</th>
<th>Error measures in the validation period</th>
<th>Qualitative considerations</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Root mean square error (RMSE)</td>
<td>• Residual autocorrelation and cross correlation plots</td>
<td>• Root mean squared forecast error (RMSFE)</td>
<td>• The appearance of forecast plots, intuitive reasonableness of the coefficients and the simplicity of the model.</td>
</tr>
<tr>
<td>• Bayesian information criterion (BIC)</td>
<td>• Durbin-Watson statistic (serial correlation test);</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Akaike’s information criterion (AIC)</td>
<td>• Non-normality test: skewness / kurtosis; Shapiro-Wilk W statistic</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Heteroscedasticity test: Breusch-Pagan test / White’s test</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Misspecification test: Ramsey RESET test</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
5. Estimation, goodness-of-fit and forecast

5.1. Correspondence traffic | Time series

Log transformation
ln_correspondence

1st differences
D_ln_correspondence

Seasonal differences (period 4)
DS_ln_correspondence

Time series: From 1Q2005 to 2Q2018 (54 observations)
5.1. Correspondence traffic | Estimation

**ARIMA MODELS**

1. **SARIMA (p,d,q) (P,D,Q)s:**
   \[\Phi(B^s)\theta(B^s)\phi(B)\Theta(B)\phi(B)\theta(B)\varepsilon_t\]
   - SARIMA (0,1,1) (0,1,1)_s

2. **SARIMA with exogenous variable:**
   \[\Phi(B^s)\theta(B^s)\phi(B)\Theta(B)\phi(B)\theta(B)\varepsilon_t\]
   - SARIMAX (2,1,1) (0,1,0) s | \(X_t = \ln \text{GDP}\)

**DECOMPOSITION MODELS**

3. **Holt-Winters' Multiplicative**
   \[L_t = a - \frac{\gamma Y_t}{S_{t-4}} + (1 - \alpha)(L_{t-4} + b_{t-4})\]
   \[b_t = b_{t-4} + \beta(L_{t-4} - L_{t-1}) + (1 - \beta)b_{t-1}\]
   \[S_t = \gamma S_{t-4} + (1 - \gamma)S_{t-1}\]
   forecast \(F_{h+k} = (L_0 + kb_0 S_{h+k})\)

4. **SARIMAX:**
   \[Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \cdots + \beta_p X_{pt} + \varepsilon_t\]

**MULTIPLE LINEAR REGRESSION**

5. **Without exogenous variable:**
   - Linear trend
   - Seasonal dummies: Q1, Q2, Q3
   - Structural breaks: D4Q2007, D4Q2011 - linear effect

6. **With exogenous variable:**
   - Exogenous variable: \(\ln(\text{PenetrationBB})\)
   - Seasonal dummies: Q1, Q2, Q3
   - Structural breaks: D4Q2011
   - Interaction regressors

---

5.1. Correspondence traffic | Goodness-of-fit evaluation

<table>
<thead>
<tr>
<th>(1.1) SARIMA</th>
<th>(1.2) SARIMAX</th>
<th>(1.3) Holt-Winters' Multiplicative</th>
<th>(1.4) MLR</th>
<th>(1.5) MLR with ln(PenetrationBB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE</td>
<td>25.54</td>
<td>25.25</td>
<td>23.01</td>
<td>23.26</td>
</tr>
<tr>
<td>AIC</td>
<td>-213.72</td>
<td>-217.62</td>
<td>(*)</td>
<td>-243.32</td>
</tr>
<tr>
<td>BIC</td>
<td>-206.16</td>
<td>-203.69</td>
<td>(*)</td>
<td>-229.39</td>
</tr>
<tr>
<td>Out-of-sample</td>
<td>24.46</td>
<td>25.16</td>
<td>20.31</td>
<td>20.95</td>
</tr>
</tbody>
</table>

Notes:
- (*) not comparable indicators;
- (**) Comparison between predicted values to real values of the time series. The set forecasts to start of 1st quarter 2018 to the end of 2nd quarter 2018.
5.1. Correspondence traffic | Forecast

Correspondence traffic decrease:
- 7% in 3Q2018
- 5% in 4Q2018
(both from the previous year)

Absolute error around 5 p.p.

Unit: Millions of objects.
Source: Authors, with data from ANACOM.
Note: Lower confidence limit (LCL) and upper confidence limits (UCL) for 95% confidence interval.

5.2. Parcels traffic | Time series

Log transformation
\( \ln_{ \text{parcels} } \)

1st differences
\( D_{\ln_{\text{parcels}}} \)

Seasonal differences (period 4)
\( DS_{\ln_{\text{parcels}}} \)

Time series: From 1Q2007 to 2Q2018 (46 observations)
5.2. Parcels traffic | Estimation

ARIMA MODELS

(1.1) ARIMA (p,d,q):
\[ \phi(B)^d y_t = \theta(B) \epsilon_t \]
ARIMA (4,1,0)

(1.2) ARIMAX: with exogenous variable
\[ \phi(B)^d \psi(B)^s y_t = \Psi(B) X_t + \theta(B) \epsilon_t \]
ARIMAX (0,1,0)(1,0,1) | X, lnPenetrationBB

DECOMPOSITION MODELS

(1.3) Holt-Winters’ Additive
\[
\begin{align*}
\text{level } & \quad L_t = \alpha(y_t - s_{t-d}) + (1 - \alpha)(L_{t-1} + b_{t-1}) \\
\text{trend } & \quad b_t = \beta(L_t - L_{t-d}) + (1 - \beta)b_{t-1} \\
\text{seasonal } & \quad S_t = \gamma(y_t - L_t) + (1 - \gamma)S_{t-d} \\
\text{forecast } & \quad F_{t+h} = L_t + hb_t + S_{t-h} \\
\end{align*}
\]

MULTIPLE LINEAR REGRESSION

(1.4) Without exogenous variable:
- Linear trend:
- Seasonal dummies: Q2, Q3, Q4
- Structural breaks: D4Q2012, D1Q2014
- Interaction regressors

(1.5) With exogenous variable:
- Exogenous variable: ln(penetrationBB)
- Seasonal dummies: Q4
- Structural breaks: D4Q2012, D1Q2014
- Interaction regressors

5.2. Parcels traffic | Goodness-of-fit evaluation

<table>
<thead>
<tr>
<th>(1.1) ARIMA</th>
<th>(1.2) ARIMAX</th>
<th>(1.3) Holt-Winters’ Additive</th>
<th>(1.4) MLR</th>
<th>(1.5) MLR with ln(PenetrationBB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE</td>
<td>61.69</td>
<td>57.70</td>
<td>50.27</td>
<td>41.32</td>
</tr>
<tr>
<td>AIC</td>
<td>-116.9</td>
<td>-121.8</td>
<td>(*)</td>
<td>-150.61</td>
</tr>
<tr>
<td>BIC</td>
<td>-109.7</td>
<td>-112.8</td>
<td>(*)</td>
<td>-139.64</td>
</tr>
<tr>
<td>Out-of-sample (RMSFE)**</td>
<td>20.55</td>
<td>27.38</td>
<td>59.24</td>
<td>30.75</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>36.59</td>
<td></td>
</tr>
</tbody>
</table>

Notes: (*) not comparable indicators;
( **) Comparison between predicted values to real values of the time series. The set forecasts to start of 1st quarter 2018 to the end of 2nd quarter 2018.
5.2. Parcels traffic | Forecast

- **Parcels traffic**: Null variation (0%) from the previous year.
- **Around 10 million objects**
- **Absolute error around 3 p.p.** (both from the previous year)

Unit: Millions of objects.
Source: Authors, with data from ANACOM.
Note: Lower confidence limits (LCL) and upper confidence limits (UCL) for 95% confidence interval.

6. Conclusion
1 IT sector has two different effects on postal sector – showed by Portuguese data

- Digitalization (e-government, e-substitution and e-invoice) has a negative impact on the correspondence postal traffic.
- E-commerce helped the parcels traffic to grow, due to the delivery of physical product brought through the Internet.

2 GDP is gaining importance again to explain postal traffic

- GDP lost its force to explain these series, mainly due the financial crisis. In recent years, it is gaining importance again, especially in the correspondence postal traffic.
- However, it has a negative effect on correspondence, in contrast with what happened in the past.
The decrease of the correspondence traffic is not expected to slow down soon, in Portugal

- Between 1Q2005 and 2Q2018, the best fitted model is Multiplicative Holt-Winters
- Forecasts show a decrease of correspondence traffic:
  - around 7% in 3Q2018 (from the previous year)
  - around 5% in 4Q2018 (from the previous year)
  - with an absolute error around 5 percentual points

The forecast parcels traffic shows a stabilization in Portugal, but may increase in the future, with the influence of other variables

- Between 1Q2007 and 2Q2018, the best fitted model is the Multiple Linear Regression, with a trend, seasonality dummies, a structural break dummy and no exogenous variables
- Forecast shows a stabilization around 10 million objects, with an absolute error around 3 percentual points (from the previous year)
Thank you!